COMPRESSED SENSING – A LOOK BEYOND LINEAR PROGRAMMING

Christian R. Berger, Javier Areta, Krishna Pattipati and Peter Willett

Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269

ABSTRACT

Recently, significant attention in compressed sensing has been focused on Basis Pursuit, exchanging the cardinality operator with the l_1-norm, which leads to a linear formulation. Here, we want to look beyond using the l_1-norm in two ways: investigating non-linear solutions of higher complexity, but closer to the original problem for one, and improving known low complexity solutions based on Matching Pursuit using rollout concepts. Our simulation results concur with previous findings that once $\mathbf{x}$ is "sparse enough", many algorithms find the correct solution, but for averagely sparse problems we find that the l_1-norm often does not converge to the correct solution – in fact being outperformed by Matching Pursuit based algorithms at lower complexity. The non-linear algorithm we suggest has increased complexity, but shows superior performance in this setting.

Index Terms— Compressed sensing, sparse estimation, non-linear programming, rollout.

1. INTRODUCTION

In sparse estimation, a signal $\mathbf{x} \in \mathbb{C}^n$ is estimated from a limited set of measurements. These measurements are linear projections of the form $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^m$, onto a known set of, generally dependent, vectors in A. The interesting aspect of this problem appears when $m < n$ and one is interested in the minimum number of measurements necessary to reconstruct the original signal. This problem only has a well-defined solution, when we can assume the signal $\mathbf{x}$ to be sparse in nature, i.e., the number of non-zero elements of $\mathbf{x}$, $k = \text{card}(\mathbf{x})$, is small – necessarily smaller than the number of measurements. Since we can't know which elements of $\mathbf{x}$ are non-zero, this generally leads to an intractable combinatorial optimization problem.

Different versions of *Matching Pursuit* had been applied in signal processing as greedy solutions, see e.g. [1, 2]; lately significant research has been focused on *Basis Pursuit*, see e.g. [3, 4, 5] and references therein. The new focus of this research, besides feasible complexity solutions, is the theoretical relationship between the minimum number of needed measurements, the sparseness of the solution and the properties of the measurement matrix A. The interest in *Basis Pursuit* was triggered by the observation that replacing the explicit sparseness constraint on $\mathbf{x}$ with its l_1-norm often leads to the same solution. This in turn enables a convex problem formulation, which can be solved efficiently using interior point methods, see e.g. [6].

Although there are some established results characterizing when the true sparseness constraint can be replaced with the l_1-norm [3], they are not tight in the sense that in many cases not covered by the cited results, the solution can still be found. Also interestingly, the condition developed in [3] applies equally to greedy algorithms (Orthogonal Matching Pursuit) as well as to the l_1-norm formulation. This gives rise to the impression that once the problem is "sparse-enough", the solution is fairly robust and can be found by different algorithms.

We in this work want to focus on algorithms that push the frontier of solvable problems, i.e., we are interested in algorithms that can find the optimal solution with a minimum number of measurements for only averagely sparse problems. We want to compare different algorithms, reducing the number of measurements and/or the degree of sparseness in $\mathbf{x}$, to find the threshold when each algorithm "breaks" from the optimal solution. In this context, we are interested in two types of algorithms:

1. We want to find a problem formulation, which replaces the sparseness $\text{card}(\mathbf{x})$ with an approximation different from the l_1-norm, leading to a closer approximation of the original problem.

2. We want to improve existing greedy algorithms based on *Matching Pursuit*.

The rest of this paper is organized as follows: in Section 2, we present our new solutions; in Section 3, we discuss implementation and numerical results. Finally, we conclude the paper in Section 4.

2. PROPOSED SOLUTIONS

2.1. Problem Formulation

We want to solve the following problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \text{card}(\mathbf{x}) \quad \text{subj. to } |A\mathbf{x} - \mathbf{y}|^2 \leq \epsilon \qquad (1)$$

Supported by the Office of Naval Research grant N00014-07-1-0429.

where $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{y} \in \mathbb{C}^m$, $A \in \mathbb{C}^{m\times n}$ and $m < n$. The uniqueness of the solution is connected to $k := \text{card}(\mathbf{x}) \ll n$, i.e., the sparseness condition on $\mathbf{x}$.

2.2. Implementation via Lagrangian Relaxation

The difficulty in the original problem is that the cardinality operator

$$\text{card}(\mathbf{x}) = \sum_{i=1}^{n} \text{card}(x_i), \quad \text{card}(x_i) = \begin{cases} 1, & x_i \neq 0 \\ 0, & x_i = 0 \end{cases} \tag{2}$$

is non-differentiable. The interpretation of the cardinality operator as a "zero-norm" is one approach,

$$\lim_{p\to 0} |\mathbf{x}|_p = \lim_{p\to 0} \sum_{i=1}^{n} |x_i|^p = \text{card}(\mathbf{x}), \tag{3}$$

where $p = 1$, i.e., the l_1-norm, is the norm with the smallest p which is convex.

We formulate the problem by using a new constraint:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x},\mathbf{z}} \sum_{i=1}^{n} z_i \tag{4}$$

$$\text{subj. to } \; x_i(1 - z_i) = 0 \; \forall i \tag{5}$$

$$|A\mathbf{x} - \mathbf{y}|^2 < \epsilon \tag{6}$$

where $z_i \in \{0, 1\}$. Applying Lagrangian relaxation, we get

$$\begin{aligned} \min J(\mathbf{x}, \mathbf{z}) &= \min_{\mathbf{x},\mathbf{z}} \sum_{i=1}^{n} z_i + \sum_{i=1}^{n} \lambda_i x_i (1 - z_i) \\ &= \min_{\mathbf{x}} \left[\min_{\mathbf{z}} \left(\sum_{i=1}^{n} z_i (1 - \lambda_i x_i) \right) + \sum_{i=1}^{n} \lambda_i x_i \right]. \end{aligned} \tag{7}$$

Evidently the minimization over $\mathbf{z}$ yields

$$z_i = \begin{cases} 1, & \lambda_i x_i > 1 \\ 0, & \lambda_i x_i \leq 1 \end{cases} \tag{8}$$

Inserting this solution back into (7)

$$\begin{aligned} \min \tilde{J}(\mathbf{x}, \boldsymbol{\lambda}) &= \min_{\mathbf{x}} \left[\sum_{i=1}^{n} \min\{1 - \lambda_i x_i, 0\} + \sum_{i=1}^{n} \lambda_i x_i \right] \\ &= \min_{\mathbf{x}} \sum_{i=1}^{n} \min\{1, \lambda_i x_i\} \end{aligned} \tag{9}$$

where now we have an additional maximization over $\boldsymbol{\lambda}$. Examining the plot of this objective function in Fig 1, we can note several things:

- If we maximize over λ, this function is equivalent to the cardinality operator.
- For $\lambda = \text{sign}(x)$, the l_1-norm is the convex extension of this function.

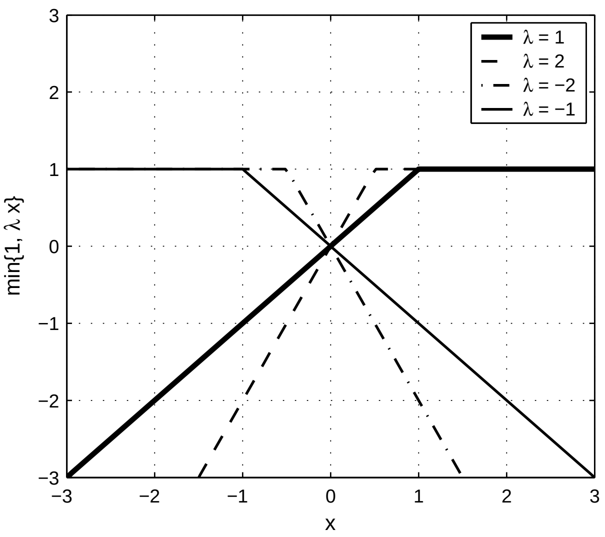

Fig. 1. The plot of $\min\{1, \lambda x\}$ shows that with concurrent maximization over λ, this objective function is equivalent to the cardinality operator.

2.3. Approximating the Cardinality Operator

The objective function in (9) is still not completely differentiable and we have to update a vector of Lagrangian multipliers $\boldsymbol{\lambda}$. Instead we suggest a continuously differentiable approximation in the form of the hyperbolic tangent,

$$J_c(\mathbf{x}) = \sum_{i=1}^{n} \tanh\left(c\,|x_i|^2\right). \tag{10}$$

In Fig. 2 we can see that as $c \to \infty$, this function converges to the cardinality operator. Use of the new objective function J_c leads to the following problem formulation consisting of a series of non-linear optimization problems:

$$\lim_{c\to\infty} \min_{\mathbf{x}} J_c(\mathbf{x}) \quad \text{subj. to } |A\mathbf{x} - \mathbf{y}|^2 \leq \epsilon. \tag{11}$$

We have the following derivatives of J_c,

$$\frac{\partial J_c}{\partial x_k} = \frac{2cx_k}{\left[\cosh\left(c\,|x_k|^2\right)\right]^2} \tag{12}$$

$$\frac{\partial^2 J_c}{\partial x_k \partial x_l} = \frac{2c\left[1 - 4c|x_k|^2 \tanh\left(c\,|x_k|^2\right)\right]}{\left[\cosh\left(c\,|x_k|^2\right)\right]^2} \delta_{kl} \tag{13}$$

where δ_{kl} is the Kronecker delta. Accordingly the gradient is well defined and the Hessian matrix has a diagonal structure. The Hessian is not positive definite, but due to the diagonal structure we can easily approximate it with a positive definite matrix to calculate a pseudo-Newton direction.

For each c this non-linear optimization problem converges to a well-defined (local) maximum. To solve, we use standard

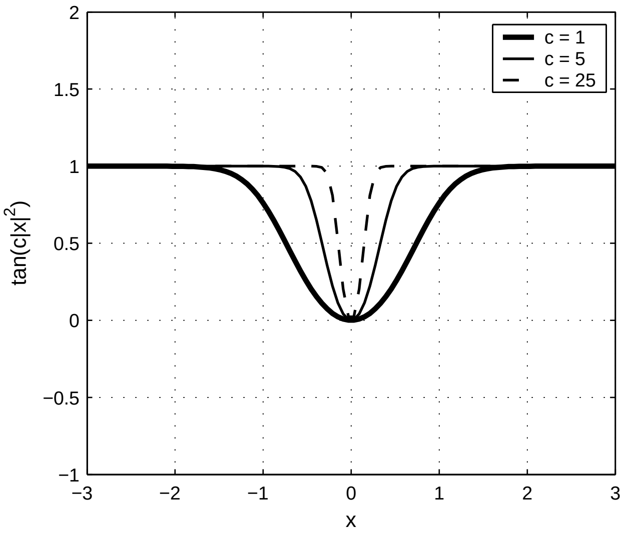

Fig. 2. $\tanh(c|x|^2)$ is a continiously differentiable function, which converges to the cardinality operator, when $c \to \infty$.

methods of non-linear programming [7]. We start with a feasible solution, e.g., the least-squares solution, and solve for some small c initially. When increasing c, we use the previous solution for initializing the new problem; and if we increase c slowly, the algorithm converges to a final solution. In practice, we will not solve each subproblem exactly, but only execute a few iterations, possibly just one, to save computation.

2.4. Improving Greedy Solutions with Rollout Concepts

Rollout algorithms were first proposed for the approximate solution of dynamic programming recursions by Bertsekas et al. in [7]. They are a class of suboptimal solution methods inspired by the policy iteration of dynamic programming and the approximate policy iteration of neuro-dynamic programming. The rollout algorithm, combined with base heuristics (e.g., Basis Pursuit,...), can solve combinatorial optimization problems such as that here with a higher computational efficiency than the optimal strategies, while being superior to those using the base algorithms only.

2.5. Reducing Complexity of Greedy Algorithms

The Order-Recursive LS MP algorithm is a variant of Matching Pursuit (MP), where the next greedy choice is based on reducing the actual fitting error. This is a sensible metric, but leads to high complexity. To reduce this complexity, we suggest an efficient implementation. In the case of orthogonal columns, maximizing the reduction in the fitting error is equivalent to choosing the maximum projection as in the usual MP algorithm. Therefore we suggest calculating orthogonal basis vectors of the space spanned by the selected columns of A, via the Gram-Schmidt procedure. When a new column is selected, we subtract its component from the columns so far not selected. This way we can add one orthogonal vector to the basis at each iteration without recalculating the previous basis-vectors.

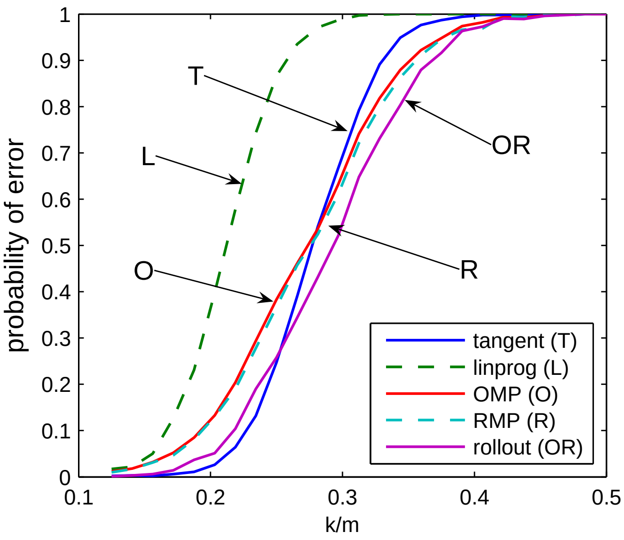

Fig. 3. Probability of error detecting the non-zero components of $\mathbf{x}$; 10^3 simulation runs, $n = 512$, $m = 64$.

3. IMPLEMENTATION AND NUMERICAL RESULTS

3.1. Implementation

Both problem formulations in Section 2.2 and Section 2.3 are highly non-linear in nature. As implementation we apply efficient algorithms based on descent directions [7], namely Trust-Region and Preconditioned Conjugate Gradient methods. The differentiable approximation of the cardinality operator in Section 2.3 has considerably lower complexity, stemming from the diagonal Hessian matrix.

For the following simulations we used available off-the-shelf products, i.e., an implementation of the Trust-Region method available in the MATLAB Optimization Toolbox, which ran in a loop updating c at each iteration and checking convergence via

$$\sum_{i=1}^{n} \left[\tanh\left(c\,|x_i|^2\right)\left(1 - \tanh\left(c\,|x_i|^2\right)\right)\right] \leq n\epsilon, \qquad (14)$$

which terminates once the approximation of the cardinality operator has converged for each x_i.

The rollout implementation used the Orthogonal MP (OMP) algorithm to find a greedy solution and then varied it by recalculating a solution while excluding one of the elements picked early by the first greedy solution. To get a noticeable improvement in performance, we found about twenty iterations to be sufficient.

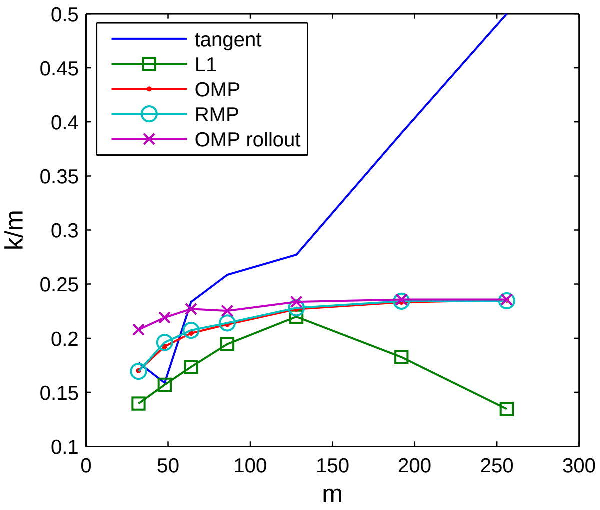

Fig. 4. Needed sparseness; plotted is the maximum ratio k/m for which each algorithm can still find the right non-zero values of $\mathbf{x}$ in more than 85 % of the cases, $n = 512$.

3.2. Numerical Results

We will compare the non-linear optimization based on the tanh function with an l_1-norm implementation, Orthogonal Matching Pursuit (OMP), order-recursive LS MP (RMP) and the OMP with rollout. We use only noiseless observations, therefore $\epsilon = 0$ in (1). We generate the signal $\mathbf{x}$ by randomly choosing k elements and assign values using a zero-mean, unit variance Gaussian distribution. The measurement matrix A has random entries from the same distribution and we normalize the columns.

As a first scenario, we fix $n = 512$, $m = 64$ and slowly increase the ratio of non-zero elements to observations from $1/8$ to $1/2$; we observe in Fig. 4 that for small k/m, i.e., very sparse problems, basically all algorithms chose the correct non-zero elements of $\mathbf{x}$. Still, different algorithms can handle less sparse problems. Especially the l_1-norm algorithm introduces many additional small elements into the solution, which can not be discerned from correct non-zero elements in the averagely sparse case, since each element of $\mathbf{x}$ is rather small itself.

Next we increase m, while keeping $n = 512$; for each m we increase the ratio k/m until the probability of error exceeds 15 %. Fig. 4 confirms the trend observed before that the l_1-norm cannot handle averagely sparse problems. This is even more so when the ratio m/n approaches $1/2$, as this further reduces the sparseness. Comparatively, even the different OMP versions perform better with the rollout outperforming RMP. The throughout strongest performance has the non-linear formulation using the tanh, but at the cost of high complexity, c.f. Fig. 5. Using our implementations all algorithms show a similar scaling behavior, which doesn't seem to give l_1-norm any edge over the OMP variations.

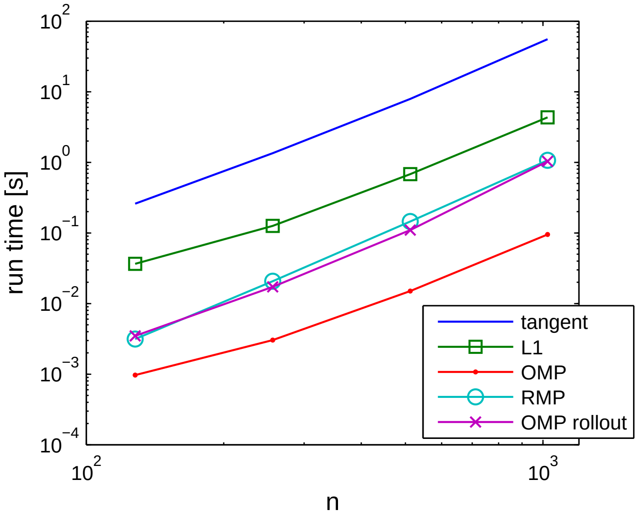

Fig. 5. Run time comparison for the different algorithms for increasing n.

4. CONCLUSION

We suggested several new approaches to the sparse estimation problem. For one, approximating the cardinality operator via the tanh function, we find a non-linear algorithm which can also solve moderately sparse problems. Second, we improve known low complexity greedy algorithms using rollout techniques, which leads to an algorithm of lower complexity than Basis Pursuit and better performance than regular OMP.

5. REFERENCES

[1] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[2] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 101–111, Jan. 2003.

[3] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] E.J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[6] S.-J. Kim, K. Koh, M. Lustig, S. Boyd and D. Gorinevsky, "A method for large-scale l_1-regularized least squares problems with applications in signal processing and statistics," 2007, Preprint, available at http://www.stanford.edu/˜boyd/reports/l1_ls.pdf.

[7] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, USA, 2 edition, 1995.